

Monitoring change in biodiversity through composite indices

S.T. Buckland^{1,*}, A.E. Magurran², R.E. Green³ and R.M. Fewster⁴

¹*Centre for Research into Ecological and Environmental Modelling, University of St Andrews,
The Observatory, Buchanan Gardens, St Andrews KY16 9LZ, UK*

²*School of Biology, Gatty Marine Laboratory, University of St Andrews, St Andrews KY16 8LB, UK*

³*Royal Society for the Protection of Birds and Conservation Biology Group, Department of Zoology,
University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK*

⁴*Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand*

The need to monitor trends in biodiversity raises many technical issues. What are the features of a good biodiversity index? How should trends in abundance of individual species be estimated? How should composite indices, possibly spanning very diverse taxa, be formed? At what spatial scale should composite indices be applied? How might change-points—points at which the underlying trend changes—be identified? We address some of the technical issues underlying composite indices, including survey design, weighting of the constituent indices, identification of change-points and estimation of spatially varying time trends. We suggest some criteria that biodiversity measures for use in monitoring surveys should satisfy, and we discuss the problems of implementing rigorous methods. We illustrate the properties of different composite indices using UK farmland bird data. We conclude that no single index can capture all aspects of biodiversity change, but that a modified Shannon index and the geometric mean of relative abundance have useful properties.

Keywords: biodiversity monitoring; indices of trend; Shannon index; Simpson's index

1. INTRODUCTION

There is little prospect of effective action to limit biodiversity loss unless biodiversity can be measured and its rate of change quantified. This need raises many technical and philosophical issues, not least because biodiversity is a concept with multiple meanings, and with attributes that can be measured in many different ways. In this paper, we assume that biodiversity will be monitored by surveying changes in abundance of individual species, and we consider how to design and analyse such surveys when we wish to combine trends across species to form composite indices.

Yoccoz *et al.* (2001) note that many programmes for monitoring biodiversity inadequately address three questions: Why monitor? What should be monitored? How should monitoring be carried out?

We assume that the primary question underlying monitoring is to ask whether biodiversity is changing over time, and if so, at what rate. This is consistent with the stated aim of the 2010 Biodiversity Target of the Convention on Biological Diversity: '...to achieve... a significant reduction of the current rate of biodiversity loss...'. We further assume that the question needs to be asked separately for different geographic regions and species groupings, but accept that it would be advantageous to combine information across species groups and regions if an index is to be comprehensible and useful to decision makers.

* Author for correspondence (steve@mcs.st-and.ac.uk)

One contribution of 19 to a Discussion Meeting Issue 'Beyond extinction rates: monitoring wild nature for the 2010 target'.

The question of which species to monitor is difficult because what is desirable is compromised by what is possible. Because different methods are required to survey different types of species (Southwood & Henderson 2000), we assume that survey programmes will be targeted at certain groups of species. Further, we assume that these programmes will generate, where possible, estimates of animal or plant density available on an annual basis (although the methods are easily applied to other time series of estimates, whether regularly spaced in time or not).

In this paper, we principally address the third of the questions posed by Yoccoz *et al.* (2001), although it cannot be taken entirely independently from the other two. In §2, we suggest criteria that a biodiversity measure should satisfy if it is to be used to quantify change over time. We consider issues related to survey design in §3, and how to measure biodiversity and changes in biodiversity in §4. Some of the methods we propose are illustrated in §5 and we discuss some of the problematic issues related to monitoring biodiversity in §6.

2. CRITERIA FOR A BIODIVERSITY MEASURE

We suggest appropriate criteria for a biodiversity measure when that measure is to be used primarily to assess changes in biodiversity over time. This provides an objective means of choosing between possible measures. We assume that three aspects of biodiversity are of primary interest: number of species, overall abundance, and species evenness (high evenness occurs

when many species have similar abundance, with no single species dominating). We further assume that our measure is across a group of similar species (e.g. a family or guild) only. If species within a group vary appreciably in size, it may be better to substitute 'biomass' for 'abundance' in the following. Species might also be weighted in different ways. For example, endemics or species of particular conservation interest could be assigned higher value. Similarly, taxonomic status or phylogenetic difference among species could be taken into account as happens in existing measures of taxonomic distinctness and phylogenetic diversity (see Warwick & Clarke 2001; Faith 2002).

1. For a system that has a constant number of species, overall abundance and species evenness, but with varying abundance of individual species, the index should show no trend.
2. If overall abundance is decreasing, but number of species and species evenness are constant, the index should decrease.
3. If species evenness is decreasing, but number of species and overall abundance are constant, the index should decrease.
4. If number of species is decreasing, but overall abundance and species evenness are constant, the index should decrease.
5. The index should have an estimator whose expected value is not a function of sample size.
6. The estimator of the index should have good and measurable precision.

3. SURVEY DESIGN

Yoccoz *et al.* (2001) note that many monitoring programmes either ignore or deal ineffectively with two primary sources of variation in monitoring data: spatial variation and detectability.

Biodiversity, and trends in biodiversity, can vary enormously between locations (reflecting differing habitats, land uses, climates, etc.), so that monitoring programmes should be designed to take account of this spatial variation. This is especially critical if biodiversity is to be monitored at a regional or global (as distinct from site) level. Too often, monitoring programmes are conducted at unrepresentative sites (sometimes called 'sentinel sites') and conclusions generalized to the landscape as a whole. For example, the British Butterfly Monitoring Scheme (BMS; Pollard & Yates 1993) uses transects at sites chosen because they are suitable for butterflies. Such sites are often protected reserves, or atypical in other respects, and trends in abundance may be unrepresentative of what is happening in the wider countryside. Another example is the North American Breeding Bird Survey (NA BBS; Droege 1990; Link & Sauer 1997), which is conducted along roads and tracks where habitats are unlikely to be typical of the area as a whole; these habitats may be avoided by some species (Reijnen *et al.* 1995).

Until recently, the main United Kingdom breeding bird monitoring programme was the Common Birds Census (CBC; Williamson 1964) and was based on non-random sites selected by volunteer observers.

It has now been replaced by the United Kingdom's Breeding Bird Survey (UK BBS; Freeman *et al.* 2003), which is based on a stratified random sample of transects from throughout the UK. Although critics argued that volunteer observers would not wish to visit random sites, where bird densities and diversity might often be low, the scheme now has nearly 2000 contributors compared with 200 or fewer for the CBC.

Differences in biodiversity measures over time at a single location may be due to real changes or may simply reflect the fact that species were more detectable in some time-points than others, perhaps due to variable observer effort, time of year, or habitat succession affecting the ease with which species could be detected, or many other possible factors. Often no attempt is made to estimate detectability because it adds an unacceptable overhead to surveys, which often span a wide range of species. Thus, the BMS is based on (often incomplete) counts of butterflies within a strip of specified width (usually 5 m but sometimes wider), while the NA BBS is based on counts of all birds detected out to 400 m from the sample point. It is certain that only a proportion of the birds present within 400 m will be detected, even of those vocalizing during the survey, and that this proportion will vary with habitat, environmental conditions and observer skill. By contrast, the UK BBS, relying on volunteer observers, uses line transect sampling, an example of distance sampling (Buckland *et al.* 2001), to correct for detectability. Recognizing the difficulty that untrained volunteer observers may have with estimating distances, the UK BBS uses just three distance intervals for tallying detections. However, given the rate at which the price of survey lasers is dropping, it may soon be realistic to expect a volunteer, in developed countries at least, to purchase one, in which case detectability in the context of distance sampling could be measured with greater precision and lower bias.

The UK BBS is not the only large-scale monitoring programme that successfully deals with variability due to both spatial variation and detectability, although other examples are rare. Another is the Waterfowl Breeding Population and Habitat Survey, which is a spring survey of breeding waterfowl in the north-central USA and Canada. This is conducted by aerial surveys of strips, using a stratified systematic sampling survey design, and detectability is measured by surveying a proportion of the strips from the ground (where detectability is assumed to be certain) as well as from the air.

When the NA BBS, the BMS and the CBC were first established, they were ground-breaking but, as time has passed, their limitations have come to light. Too often, established but flawed methods are retained in order to avoid breaking a long and valuable time series. However, if that time series is compromised to the extent that trend estimates may seriously mislead managers, then the decision to change methods should be made. The British Trust for Ornithology faced this difficult decision when it replaced the CBC by the UK BBS. It addressed the issue of continuity of time series by running the schemes in parallel for several years to allow calibration. Freeman *et al.* (2003) found that,

despite the fact that CBC sites were non-random, for a large majority of species considered, there was no significant difference between population trends, calculated from CBC and BBS. However, these analyses were restricted to that part of the country where CBC data were sufficient to support a meaningful comparison.

Given the number of long-term time series of data from non-random sites, there will continue to be considerable interest in how best to estimate trends for wider regions within which these sites fall. Post-stratification can help reduce bias in trends. If the wider region can be divided into strata, within which trends are relatively homogeneous, then the sites that fall within each stratum might be assumed to be representative for that stratum. A common difficulty with this approach is that less-favoured habitat types may have inadequate sample sizes, yet may account for the majority of many populations. For example, some butterfly species occur at high densities within the kind of site that is monitored, and at lower densities (but higher overall abundance) through much larger tracts of less suitable habitat. Supplementing existing schemes with additional sites in under-sampled strata may be a less costly, but technically less satisfactory, option than replacing an existing scheme altogether.

Monitoring programmes should be designed so that they address the defined objectives. If practicalities lead to a design that cannot meet its objectives, then the programme should be re-examined and other options evaluated. Any programme that seriously attempts to monitor biodiversity should address the two issues of spatial variation and detectability. [Danielsen *et al.* \(2003\)](#) argue that designs are too complicated and programmes too costly for developing countries, so that simpler schemes are needed. We wholeheartedly support the response of [Yoccoz *et al.* \(2003\)](#) to this, that the 'why', 'what' and 'how' of biological monitoring is important irrespective of available resources. We also endorse the call to measure detectability in large-scale surveys from [Pollock *et al.* \(2002\)](#).

In the context of point counts (used, for example, by the NA BBS), [Buckland \(in preparation\)](#) noted that 'comparisons of counts across species are invalid, because different species have different detectabilities, and comparisons within a species across different habitats are invalid, because different habitats result in different detectabilities. Even comparisons over time in counts made at the same locations are compromised if habitat succession affects detectability, or if an observer's hearing ability changes over time, or if observers change or, in the case of surveys near roads, if traffic noise increases over time.' Observer variation in detectability has been well demonstrated for the NA BBS ([Sauer *et al.* 1994](#); [Kendall *et al.* 1996](#); [Link & Sauer 1998a,b](#)). Detectability can be safely ignored only if detection is certain (or nearly so) within the sampled plots. It would often be necessary to have very small plots or very narrow strips along transects to ensure this, in which case many potential records beyond the plot are discarded.

Our favoured strategy is to design a survey that will fully meet its objectives (assuming that those objectives

are realistic and achievable). The region of interest may span different administrative areas and possibly several nations. Within some areas, the survey design may be achievable from the outset. In others, sampling may need to be restricted to localities that are safe or accessible, sampling methods may have to be simplified and the number of sampling locations may have to be reduced (achievable without bias, using a stratified sampling scheme such as that used by the UK BBS). However, the full design should remain as a goal for those areas to aspire to. A design that may be unachievable now may well be achievable in 10 years, especially if other areas in the region are able to implement the full design successfully. If the simpler methodologies or reduced sampling rates are carefully planned, this need not compromise the long-term time series; rather, as areas acquire the expertise or resources to upgrade their part of the programme, the aim should be to make upgrades 'backwards compatible'. That is, it should be possible to extract data from the improved programme that are comparable with those from the simplified programme.

Again, we concur with [Yoccoz *et al.* \(2003\)](#) that there is no necessity for sound survey design to lead to a complex monitoring programme; a well-designed programme makes for easy data analysis, whereas a poorly designed one leads to either flawed or complex data analysis, and often both.

It is possible to use more ambitious survey designs in programmes that use a few professional observers compared with those that use a large number of volunteers. For the latter, it is difficult to get the balance between methods that are over-simplistic and methods that are complex to the point that compromises both data quality and the goodwill of the volunteers. However, if field methods are simplified to the point that the data cannot possibly answer the objectives of the programme, then the survey fails the volunteers, who have contributed their effort to help achieve those objectives. Moreover, a combination of different sampling techniques may be required to produce an accurate representation of biodiversity. For example, [Sørensen *et al.* \(2002\)](#) used six different sampling methods in their investigation of spider diversity in Tanzania. Finally, the challenges of identifying less charismatic taxa—including most invertebrate groups—may impede a comprehensive monitoring programme.

4. METHODS OF MEASURING BIODIVERSITY

(a) *General measures*

[Magurran \(2004\)](#) gives a comprehensive account of methods for measuring biodiversity. She defines biodiversity to be 'the *variety* and *abundance* of species in a defined unit of study' (emphasis added). Thus, there are two concepts here. Abundance is easily defined as total numbers of individuals or the density of individuals, though biomass or percentage ground cover (for terrestrial plants) may also be appropriate measures. Note that, if we choose to measure abundance as biomass rather than number of individuals, we may observe very different trends over time if we

combine our measure across species. Variety is less easily defined. Hence, it is unrealistic to expect to encapsulate biodiversity in a single measure, or indeed in just two.

There are many databases around the world that comprise recorded presence by species and locality. These are considered by many to be a valuable resource from which changes in biodiversity may be inferred. We contend that reliable quantification of trend in biodiversity is only possible through well-designed and coordinated surveys for the simple reason that biased or incomparable samples can produce misleading conclusions. For example, databases that record presence of species tend to give a falsely optimistic view of recent changes, owing to the general trend towards greater participation in natural history recording. Even if this increase in effort is modelled, different models can give rise to very different estimates of trend in biodiversity.

We do not favour the use of counts of number of species ('species richness') for monitoring changes in biodiversity. Trends in such counts are prone to bias because detectability changes over time; changes in observer effort may make a significant difference, for example. However, in some cases, it may not be feasible to estimate abundance, for example in some invertebrate surveys, in which case the number of species within a standard sampled quadrat ('species density'; Magurran 2004, p. 75) might be recorded. Like species richness, species density measures may be biased because of changes in detectability over time if the quadrat is too large to ensure certain detection. This problem can be addressed by estimating detectability in a rigorous way. Often an easier solution is to use a small sampling unit, for which a species is certain to be detected if present. For example, for small plants, this might be a 1 m² quadrat. (Species richness tends to be measured for much larger units.) For comparability, it is important that the quadrats are a standard size. Simply converting species richness to species density by dividing by plot area when that area varies does not provide a valid measure of species density, because the number of species is expected to increase nonlinearly with area (Gotelli & Colwell 2001).

An approach that avoids some of the drawbacks of species richness and species density is to select random quadrats as above, and to record the presence/absence of each species by quadrat. Instead of converting these data to species density, we can use the number (or proportion) of quadrats occupied by a species (occupancy) as an index of its abundance. Such indices might then be combined across species into a composite index (see also §6). Note, though, that the relationship between occupancy and abundance is nonlinear (Seber 1982; Thompson *et al.* 1998, pp. 78–79).

Another difficulty with species richness and species density is that they depend heavily on the size of the recording unit (site or quadrat), and trends are not robust to the choice of size. Hence, even if we decide to use a fixed size of quadrat, our trend estimate will depend on our choice of size. Species density is limited as a measure of biodiversity change because it does not

capture information on changes at large spatial scales and this is made worse by using small quadrats (see §6).

All of the information about community change is contained within the time series of species abundance distributions. The difficulty lies in how to extract a suitable measure of change in biodiversity. In the next section, we define some potentially suitable measures. These are typically called indices and the definitions suggest that they are calculated from the true abundance distributions. Of course, in practice, they are generally estimated from sample survey data, especially when we are interested in changes in biodiversity over large regions. Note that our criteria 5 and 6 are properties of estimators of the indices, whereas criteria 1–4 apply to the indices themselves.

(b) *Specific measures*

We are primarily interested in monitoring changes in biodiversity within a wider region through time. We assume that the population density of each of a group of m animal or plant species is estimated at each of a number of sites within the region. With this objective and assumption in mind, we explore the following measures of biodiversity.

Overall density. An estimate of the number of individuals per unit area, obtained by estimating the density of each species in the group and summing across species. Defining d_{ij} to be the mean density (number of individuals per unit area) across sites of species i in year j , the index for year j is $A_j = \sum_i d_{ij}$.

Arithmetic mean of relative abundance indices. The species-specific densities d_{ij} are scaled by dividing the time series for each species by its estimated density at the initial time point. The resulting relative abundance indices are then averaged:

$$R_j = \frac{1}{m} \sum_i \frac{d_{ij}}{d_{i1}}$$

Geometric mean of relative abundance indices. As for the previous measure, except that a geometric mean of the relative abundance indices is taken. Equivalently, the indices are averaged on a log scale and the average is exponentiated:

$$G_j = \exp \left(\frac{1}{m} \sum_i \log \frac{d_{ij}}{d_{i1}} \right)$$

Simpson's index. We define $p_{ij} = d_{ij} / \sum_i d_{ij}$ to be the proportion of individuals present in year j that belong to species i . **Simpson's (1949)** index for year j is then $D_j = \sum_i p_{ij}^2$. Low values of the index correspond to high diversity. It is convenient therefore to use a transformation, such as $1/D_j$, $1 - D_j$ or $-\log_e D_j$ (Magurran 2004, p. 115). We use $-\log_e D_j$ (see below).

Shannon index. The Shannon index is $H_j = -\sum_i p_{ij} \log_e(p_{ij})$.

Modified Shannon index. Suppose we define $q_{ij} = d_{ij} / \sum_i d_{i1}$. Hence, in year 1, $q_{i1} = p_{i1}$, but in subsequent years, the q_{ij} are standardized by dividing by the sum of densities in year 1; unlike the p_{ij} , their sum for year j ($j > 1$) is not constrained to be unity.

Then we define the modified Shannon index in year j to be $M_j = -\sum_i q_{ij} \log_e(q_{ij})$.

(c) Putting measures into practice

Note that, for all of the above measures, a statistical model is likely to be needed in practice to impute densities for missing sites in any given year. In §5, we do this using a generalized additive model with a Poisson error structure and a log link function. The imputation of missing values avoids the bias arising from site turnover that affected the results of Houlahan *et al.* (2000), as noted by Alford *et al.* (2001), and the use of a log link avoids the problem of negative imputed counts from which the method of Alford *et al.* (2001) suffered, as pointed out by Houlahan *et al.* (2001). The method also avoids the problem of drift that chaining methods are subject to (Geissler & Noon 1981; ter Braak *et al.* 1994).

Consider the following example: species A decreases from an average density of 0.50 individuals ha^{-1} to 0.20 individuals ha^{-1} in 10 years. Species B decreases from 0.40 to 0.16 individuals ha^{-1} in the same period. Species C increases from 0.01 to 0.04 individuals ha^{-1} . Hence, overall density has declined from 0.91 to 0.40 individuals ha^{-1} , a 56% decline in 10 years and two of the three species have declined. Now, convert these to relative abundance indices: species A goes from 1.00 to 0.40; species B from 1.00 to 0.40; and species C from 1.00 to 4.00. Taking the arithmetic average of these, we find that biodiversity has increased from 1.00 to 1.60—a 60% increase. The use of the geometric mean ameliorates this effect: the biodiversity measure declines from 1.00 to 0.86, a 14% decline. This is intermediate between the 56% decline in our index based on overall density and the changes in Simpson's and the Shannon index (23% increase and 27% increase, respectively, reflecting the increased evenness due to increased abundance of the rare species and reduced abundance of the common species).

The above example aids understanding of what the different indices measure but gives little insight into which method is best for which purpose. We consider below how each measure fares with respect to our six criteria for monitoring biodiversity.

Overall density. This measure fully meets criteria 1, 2, 5 and 6. The measure is dominated by the common species, since species are weighted by their abundance. This yields high precision but it measures only one component of biodiversity—that is, numerical abundance with species identity being ignored. If overall abundance is constant but species evenness is decreasing and/or number of species is decreasing, the measure remains constant; thus, the measure fails criteria 3 and 4.

Arithmetic mean of relative abundance indices. Our simple example above seems to suggest that this method is unsatisfactory. If relative abundance indices are combined using an arithmetic mean, species that are increasing by a constant proportion per year carry greater weight than species that are decreasing at the same proportional rate; the averaging is on the wrong scale. Under this method, criterion 1 is not satisfied in general. Consequently, there is no guarantee that

criteria 2, 3 and 4 will be satisfied. Criterion 5 is satisfied but the impact of rare but increasing species means that the index has poor precision, violating criterion 6.

Geometric mean of relative abundance indices. By averaging on the log scale, the poor properties of the previous measure are largely averted. Crucially, criterion 1 is now satisfied, as are criteria 2 and 3. Criterion 4 is problematic as a geometric mean cannot be calculated if the relative abundance is zero but the common practice of adding a small positive constant to all counts in series with zero values may give a reasonable solution. Criterion 5 is met but there may be some detrimental effect on variance if rare species are included in the index (criterion 6).

Simpson's index. Provided the distribution of the p_{ij} remains the same over time when number of species, overall abundance and species evenness remain constant, all three forms satisfy criterion 1. They also satisfy criteria 3 and 4 (although they are not very sensitive to species richness). Note that Simpson's index shows no change if all species in the group decline at the same rate—the various forms do not satisfy criterion 2. Which form of Simpson's index is best? The reciprocal $1/D_j$ fails criterion 6 (Rosenzweig 1995), whereas $-\log_e D_j$ is the only one that satisfies criterion 5 (Rosenzweig 1995). Hence, it is perhaps the best option.

Shannon index. The Shannon index is equal to the $-\log_e D_j$ form of Simpson's index when $p_{ij}=1/S$ for all i , where S is the number of species recorded. The two indices take similar values for moderate departures from an even distribution. The Shannon index satisfies the same criteria (criteria 1, 3, 4, 5 and 6) as $-\log_e D_j$, but is considered inferior (e.g. Magurran 2004; p. 101) because it is less sensitive to shifts in the underlying distribution of species abundances (May 1975) and because it may have substantial bias, particularly when only a small proportion of species have been sampled (Lande 1996). Simpson's index also has a smaller variance (Lande 1996). The two indices will give concordant rankings when the communities involved follow approximately the same species abundance distributions. Since most assemblages tend towards a log normal distribution (May 1975; McGill 2003), we can expect both the Shannon and Simpson measures to yield similar conclusions about diversity in the majority of cases. However, environmental perturbations such as pollution or eutrophication can lead to marked changes in the relative abundances of species (Moran & Grant 1991; Magurran & Phillip 2001). This is typically manifested in an increase in dominance because only a subset of species can cope with the new conditions. In these cases, which are precisely the ones we need to identify, a dominance measure such as Simpson's index should perform better.

In circumstances where monitoring is based on a pre-determined set of species, and each species is recorded at each time-point, the shortcomings of the Shannon index are possibly of little consequence; its poor properties tend to be exposed when some species are not represented in each sample (e.g. Pla 2004).

Modified Shannon index. Neither the Shannon index nor Simpson's index reflects changes in overall abundance: if all species within a community are declining at the same rate, then both indices are stable. By contrast, our modified version of the Shannon index declines if all species in a community are declining at the same rate. Indeed, it appears to meet all six of our criteria. (Simpson's index is not so readily modified; if the q_{ij} are calculated as for the modified Shannon index, then $-\log \sum_i q_{ij}^2$ increases when all species decline at the same rate, whereas we need it to decrease.) Note that the trend in this modified index changes if the choice of baseline year changes. However, this dependence is very mild unless overall abundance changes dramatically. Note, however, that our motivation for the modified Shannon index is purely pragmatic, in the sense that we have adjusted the index so that it satisfies all of our criteria; we do not provide any theoretical underpinning.

(d) *Combining indices*

The UK wild bird index (WBI) combines trends from 139 common species using the geometric mean of relative abundance indices (Balmford *et al.* 2003; Gregory *et al.* 2003). Because rare species are excluded, the WBI has good statistical properties.

It is often argued that some species—possibly rare or endemic species—should carry greater weight. The difficulty with assigning variable weights to species (apart from the subjectivity involved) is that it becomes impossible to satisfy criterion 1, and this leads to the possibility of unpredictable and undesirable properties of any index.

Suppose we have some index I_i for species i , and we decide to assign variable weights to species so that species i has weight w_i , with $\sum w_i = 1$ and $w_i > 0$ for all i . If we take the arithmetic mean of the products $w_i I_i$, different choices of weights yield different trends in our composite indices. A mathematical property of the geometric mean of the $w_i I_i$ is that the trend is algebraically identical whatever the choice of weights w_i .

Suppose we have calculated an index for each of several groups of species. For most purposes, we would probably want to keep the indices separate, but for 'headline' purposes, we may wish to combine them. This raises the issue of weighting, especially since one index may be based on number of individuals, while another might be some relative measure of abundance. Even if both are based on number of individuals, we would not want to average (or sum) abundances (or densities) of say birds with those of spiders. In these circumstances, it seems more sensible to set subjective weights for each species group. Thus, if we have two sets of indices evaluated at time points $1, \dots, T$, say a_1, \dots, a_T and b_1, \dots, b_T , and we decide to assign weight w_a to the first and w_b to the second, with $w_a + w_b = 1$, then the composite index takes the values $w_a a_1 + w_b b_1, \dots, w_a a_T + w_b b_T$. This readily extends to say g sets of indices, where the g weights sum to one. However, given the sensitivity of the estimated trends to choice of weights, and given that these weights will inevitably be subjective with no rigorous scientific basis, a better solution would appear to be the one adopted for

the Living Planet Index: take the geometric mean of the index values at each time point. As noted above, the geometric mean of relative abundance indices is invariant to choice of weights, avoiding the need to specify subjective weights that have no rigorous scientific basis. The Shannon and Simpson's indices are not appropriate for use on relative abundance indices, or when trends from a wide variety of taxa are being combined, so that the unit of an individual animal is an inappropriate common currency.

(e) *Identifying change-points*

We may want to identify change-points in our composite index—points in time at which the rate of change in our biodiversity measure itself changes. In fact, the 2010 Biodiversity Target requires this. If we fit a smooth trend through our indices, for example using generalized additive models (Hastie & Tibshirani 1990), rate of change is measured by the slope, or first derivative, of the smooth, while change in the rate of change is measured by the second derivative. Fewster *et al.* (2000) exploited this and used the bootstrap to identify time-points at which the second derivative differed significantly from zero. As noted by Fewster *et al.*, the decision on how much to smooth the indices depends on the objectives of the analysis. Generally, interest will be in the patterns in long-term trend, rather than in short-term fluctuations. For generalized additive models, the degree of smoothing is controlled by specifying degrees of freedom; one degree of freedom corresponds to maximum smoothing (i.e. linear trend), while no smoothing at all corresponds to $T-1$ degrees of freedom, where T is the number of annual indices available in the time series. Fewster *et al.* (2000) found that setting degrees of freedom equal to about $0.3T$ proved satisfactory, although they noted that different choices should be examined before deciding on a value.

Thomas *et al.* (2004) describe methods for estimating trend and quantifying precision, treating sites as either random or fixed. When properly designed surveys are used for monitoring diversity, there is a solid basis for quantifying precision of biodiversity measures and trend estimates for a region, treating sites as random. For example, if a stratified random sampling scheme is adopted, an easy way to quantify precision is to use the non-parametric bootstrap and select resamples of locations in each stratum. The data in each of B resamples are analysed as if they were the original data. For any given year, an approximate 95% confidence interval for the biodiversity measure is obtained by ordering the B bootstrap estimates from smallest to largest and extracting the 2.5 percentiles of the distribution. If there are too few sample locations for this approach to be reliable (say 15 or fewer within a stratum), or if the sample locations are not selected according to a random scheme, a possible solution is to bootstrap species rather than sites. This would be appropriate if the species sampled are considered to be representative of all species within the community for which a biodiversity measure is required. A shortcoming of this approach is that not all recorded species will appear in any given resample; if there is a species

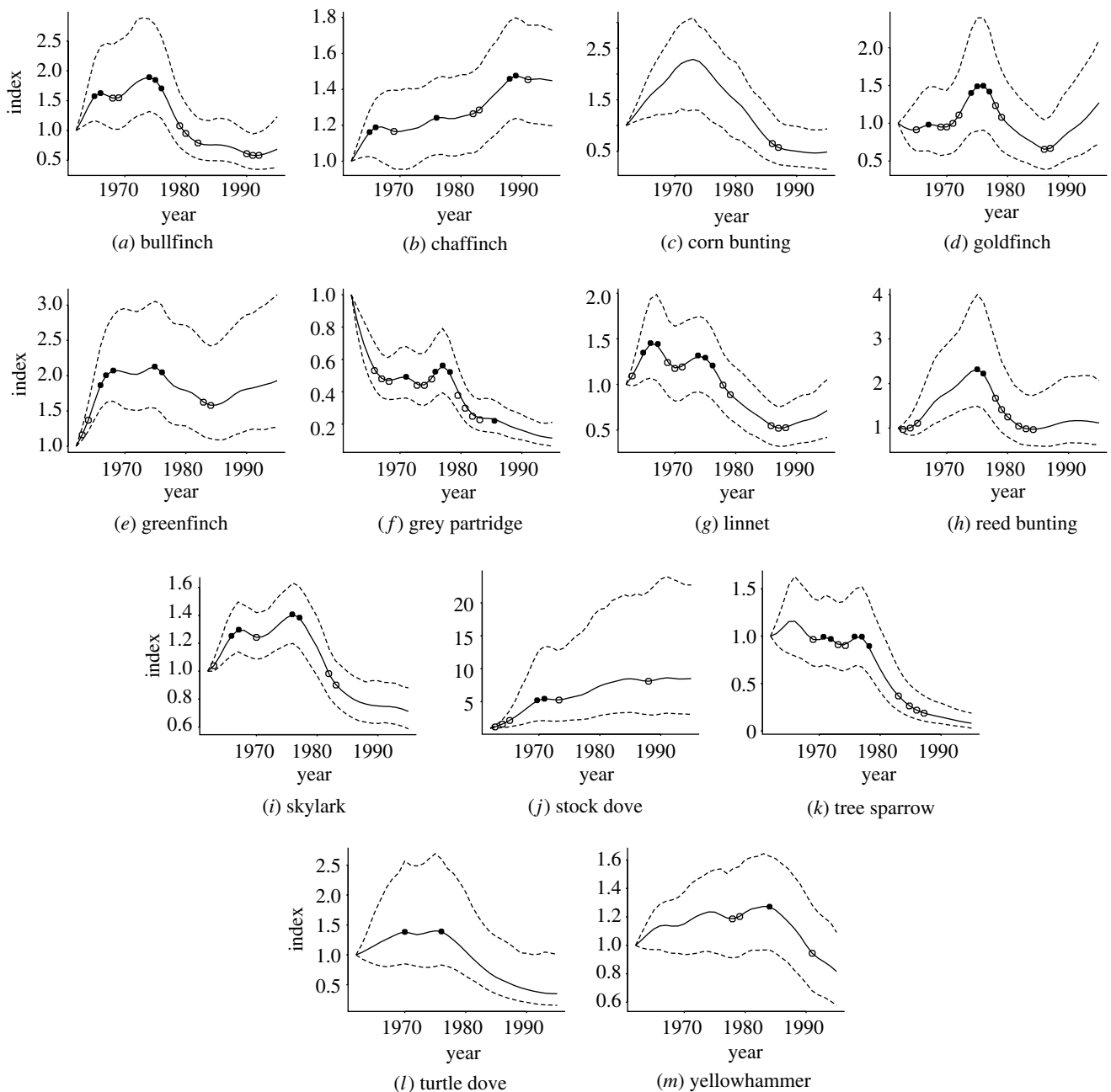


Figure 1. Trends in relative abundance of 13 farmland bird species using generalized additive models to smooth CBC counts. Also shown are 95% bootstrap confidence intervals obtained by resampling sites. The filled circles are points at which the trend showed a significant negative change; open circles indicate significant positive changes.

that dominates a community, whether or not this species is in the resample may affect the overall trend estimate appreciably, in which case this method may yield very wide confidence intervals.

5. EXAMPLE

We illustrate the various indices using the data of [Fewster *et al.* \(2000\)](#). These data are of 13 farmland bird species monitored by the CBC and the purpose of the study was to quantify evidence for adverse changes affecting many farmland species in the 1970s. [Fewster *et al.* \(2000\)](#) obtained indices for each species independently by modelling CBC counts from 1962 to 1995 using generalized additive models ([figure 1](#)). Evidence of a downturn in the 1970s was inferred from

the fact that 10 of the 13 species showed significant adverse changes between 1972 and 1978. [Siriwardena *et al.* \(1998\)](#) discuss the reasons for these changes.

In this paper, we form a composite index from the 13 separate indices by taking the geometric mean of the indices. The result is shown in [figure 2](#). We also show the composite index formed by taking the arithmetic mean. Trends for the stock dove (*Columba oenas*) were markedly different from those for other species, with approximately an eightfold increase in the index from 1962 to 1995, as the species recovered from a steep decline caused by organochlorine seed dressings, which were banned in the early 1960s. Whether or not this species is included changes inference entirely if the arithmetic mean is used, but conclusions are largely unaffected if the geometric mean is used ([figure 2](#)).

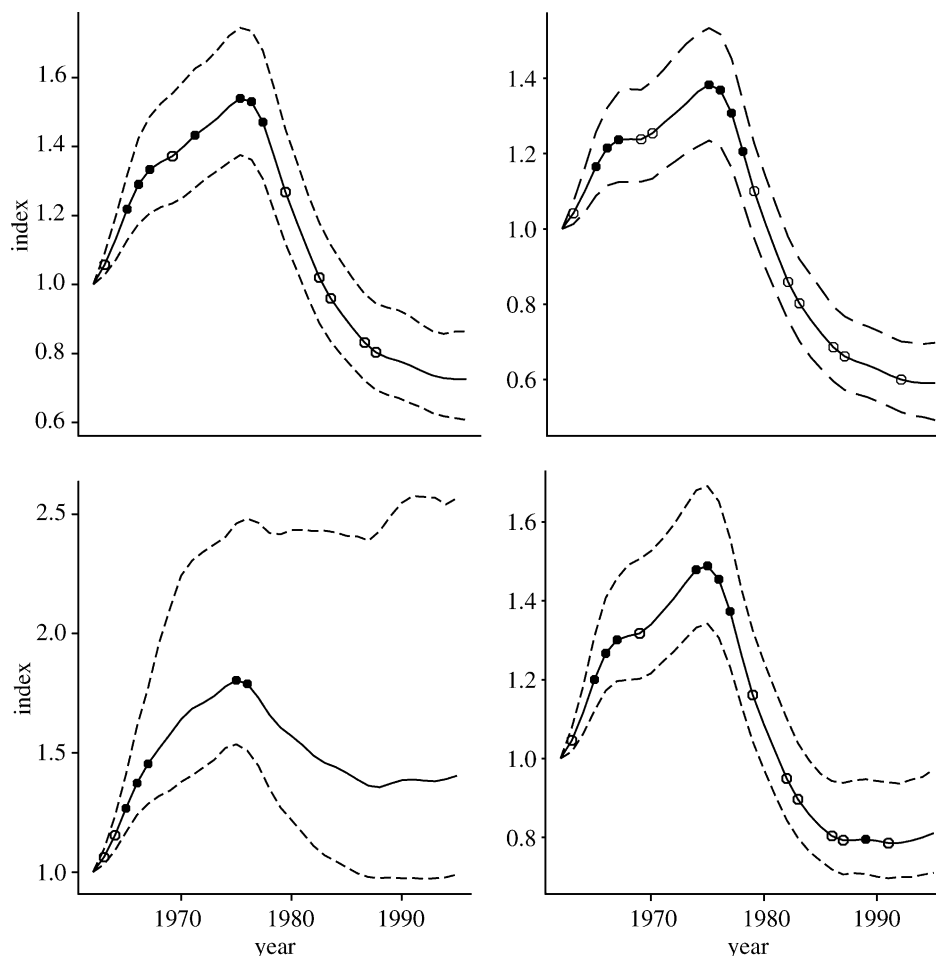


Figure 2. Trends in composite indices for biodiversity of farmland birds. The geometric mean (upper plots) and arithmetic mean (lower plots) of the smoothed trends of [figure 1](#) are shown. Plots on the left include stock dove while those on the right do not. Also shown are 95% bootstrap confidence intervals obtained by resampling sites. The filled circles are points at which the trend showed a significant negative change and the open circles indicate significant positive changes.

To assess significant changes in trend for [figure 2](#), second derivatives were evaluated for the logarithm of the curve for the composite index. For each species, 399 bootstrap resamples were generated by resampling sites and a composite index was formed for each of the 399 sets of resamples. Points at which a 95% bootstrap confidence interval for the second derivative did not include zero are indicated in [figure 2](#). These intervals were calculated using the logarithm of the curve since a constant rate of increase or decrease in abundance generates an exponential curve on the untransformed scale but a straight line on the log scale. Hence, the change-points identified are the years in which there was a change in the rate of change of abundance and this is exactly what is needed for the 2010 Biodiversity Target. If the tests are conducted on the untransformed curve, then they assess whether the absolute change in abundance with time is changing. As an example, if a population has an index value of 1 in year 1 and decreases by 50% per year, its index decreases to 0.5 in year 2 and 0.25 in year 3. By contrast, if it loses 50% of its initial population each year, the index would drop to 0.5 in year 2 and 0 (extinction) in year 3. The former corresponds to a constant rate of decline and the second to a constant absolute reduction in abundance. As noted by a referee, if tests are conducted on the

untransformed curve and biodiversity shows exponential decline, absolute declines will become smaller as biodiversity decreases so that, eventually, we will erroneously conclude that there has been a reduction in the rate of loss of biodiversity. In this context, therefore, tests should clearly be conducted on the logarithm of the curve.

The composite index using the geometric mean shows very clearly how biodiversity among farmland species changed between 1962 and 1995 ([figure 2](#)). The trend was steeply upward in the early 1960s, with a slowing down in the late 1960s. A steady increase continued until around 1975. A substantial reversal followed, with a highly significant downturn in the mid-to-late-1970s. The rapid decline slowed during the 1980s and there was a slow but steady decline from 1987 to the end of the time series. The composite index gives a much clearer picture of overall health of the British farmland bird community than do the 13 individual trends of [figure 1](#).

We show the trend lines obtained from each of the measures of §4 in [figure 3](#). The second and third plots of this figure have identical trend lines to the two left-hand plots of [figure 2](#), although the confidence intervals differ (see below).

The first plot of [figure 3](#) shows that overall density

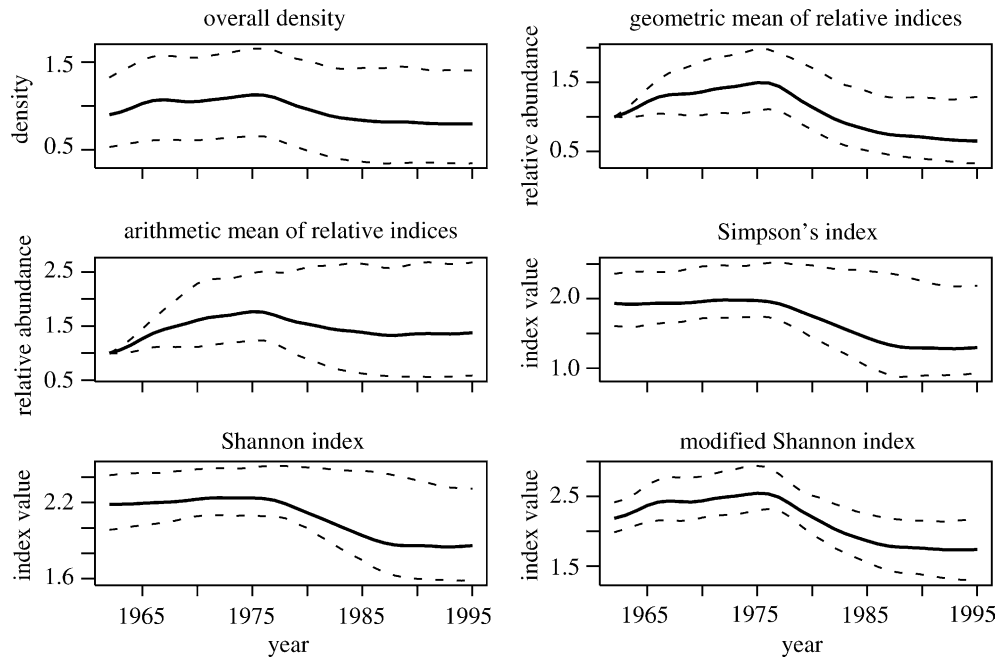


Figure 3. Trends in composite indices for biodiversity of farmland birds. The first index shows trends in overall density on CBC plots, summing across the 13 species. The second and third indices are the geometric mean and arithmetic mean of relative abundance indices and correspond to the left-hand plots of [figure 2](#). Trends in Simpson's index are shown in the fourth plot, the Shannon index in the fifth plot and the modified Shannon index in the sixth plot. Also shown are 95% bootstrap pointwise confidence intervals obtained by resampling species.

increased initially, corresponding to a recovery from the cold winter of 1962/1963. A period of stable density ended with a downturn around 1976. Densities stabilized at a lower level from around 1985. The geometric mean of relative abundance indices shows a similar pattern but the early increase and subsequent decline are greater in magnitude, suggesting that the fluctuations affected the scarcer species to a greater extent than the common species. The arithmetic mean of relative abundances shows a relatively modest decline in the late 1970s and early 1980s but, as shown in [figure 2](#), this is due to the disproportionate effect of the stock dove. The Shannon and Simpson's indices both show a decline in biodiversity between around 1975 and 1985 but, interestingly, neither show any trend in the years following the cold winter of 1962/1963. Although densities were increasing in this period, species evenness was fairly constant. The modified Shannon index shows the effect of increasing densities through this period and shows a steeper decline in biodiversity between 1975 and 1985. It is clear that both species evenness and densities declined in this later period and the modified index reflects both changes. The trends in the modified Shannon index are very similar to those in the geometric mean of relative abundance indices.

The bootstrap confidence intervals of [figures 1 and 2](#) are obtained by resampling sites and repeating the generalized additive modelling on each resample. The intervals were then obtained using the percentile method. For the case of the composite indices of [figure 2](#), this means that inference on trend in biodiversity is restricted to the 13 species analysed. From [figure 2](#), it is apparent that the decline in

biodiversity within this community of 13 species is highly significant. By contrast, the bootstrap confidence intervals of [figure 3](#) are obtained by bootstrapping species. This shows that, if we wish to draw conclusions about trends in biodiversity for a wider community for which the 13 species monitored are assumed to be representative, uncertainty is substantially greater. Indeed, only the modified Shannon index generates confidence intervals for the 1990s that do not overlap with those for the early 1970s, which suggests that it may be a more efficient indicator than the others. Note, however, that 13 species are too few for reliable quantification of variance by this means; this may explain the markedly asymmetric intervals obtained for the Shannon and Simpson's indices. Bootstrapping species gives an especially wide confidence interval for the overall density. This reflects the influence of the chaffinch, which contributed over 40% of the overall density; estimates from bootstrap resamples were therefore dominated by the number of times the chaffinch time series appeared in the resample.

6. DISCUSSION

Much ecological theory, including theory of resource management, assumes that there is a steady state, for example corresponding to carrying capacity, or a predator-prey system in balance. Reality is seldom so simple. Thus, the reason to monitor biodiversity should not be a desire to maintain current or past relative abundances of species; rather, it should be a tool for allowing decision makers to maintain biodiversity (or at least slow its loss), recognizing that some species will decline, some expand and yet others will cycle.

We believe that there should be a trend towards

common monitoring schemes spanning nations within a region or globally. Such schemes will allow for greater power in measuring changes and determining the reasons for them, and will provide an economy of scale so that nations that otherwise would have been unable to monitor their own biodiversity adequately can participate. It should be possible to enter such schemes at various levels so that nations with few resources can participate with lower sampling rates, perhaps monitoring a subset of species, possibly with simplified methods. It is essential that monitoring programmes are able to be implemented at least at some level in developing countries as it is here that biodiversity is often greatest and also most at risk. Programmes must be sufficient to meet their goals but should avoid unnecessary complexity or expenses.

Different indices measure different aspects of biodiversity. An index based on overall abundance exclusively measures a single component of biodiversity. This makes it easy to understand and interpret. However, it should be used together with an index that measures species evenness, such as the Shannon index or Simpson's index. If an index is obtained by averaging relative abundance indices across species, then the geometric mean has much better properties than the arithmetic mean. The geometric mean can also be used to form an average of composite indices or to average species-specific indices that range over a wide variety of taxa in circumstances where the species proportions required by Simpson's index and the Shannon index cannot sensibly be calculated (for example, when the indices use different units of measurement). The advantage of the latter two indices over the geometric mean of relative abundances is that both can be routinely calculated even if some species are absent in some years, whereas zero counts must be replaced by a small positive quantity to allow the geometric mean to be calculated. The modified Shannon index appears to perform very well, satisfying all six of our criteria. The trend line for the farmland birds' data set is reassuringly similar to that for the geometric mean of relative abundance indices.

If a single site with a unique habitat is to be monitored, all of the indices we consider could prove inadequate; loss of the habitat (and its associated specialist species) might lead to creation of a more common habitat and an influx of common species that exploit that habitat. As a consequence, species richness and overall abundance would both increase, as might species evenness, while the specialist species disappear. One solution is to restrict the indices to data on the specialist species of interest. By contrast, for schemes that monitor biodiversity over a wider (and heterogeneous) region, the loss of a unique habitat and its associated species within that region will reduce species evenness and species richness, so that all of our indices except the one based on overall density should detect a loss of biodiversity. This is related to the concept of β -diversity (Magurran 2004, pp. 162–184); a region will have greater biodiversity (γ -diversity) if it has diverse habitats with distinct species communities than if it is more homogeneous with a common set of species throughout.

If the objective is to monitor biodiversity across a large region that includes several habitats, the existence of β -diversity is a principal reason why measuring changes in the average species richness within small plots is unsatisfactory, even if they have wide geographical and habitat coverage. As a simple illustration, suppose that a region consists of two vegetation communities, forest and grassland. The average number of species per quadrat is the same in each but there is no overlap in species composition (high β -diversity). If grassland encroached on forest to eventually cover the entire region, the characteristic species of forest would probably all be lost, but this would not be reflected in the region-level measure of average species density, which would show no trend. To avoid the pitfalls in using a species richness measure for monitoring biodiversity, we suggested in §4 that a site might be sampled by a number of quadrats, each of the same size and small enough so that all species within a quadrat can be enumerated. If sites themselves are of a standard size (say 1 or 10 km²), are selected according to a (stratified) random scheme and each is sampled by the same number of quadrats (systematically or randomly placed), then within the wider region that the sites represent, we can use as a species-specific index of change the average number of quadrats occupied per site (with appropriate weighting if the scheme is stratified) for each species. The index can be scaled to equal unity in the initial year and the resulting species trends can be averaged using a geometric mean to estimate trend in biodiversity. Although we prefer measures based on abundance, this measure will reflect, albeit imperfectly, changes in both α -diversity and β -diversity.

Biodiversity indices are principally useful for providing empirical measures for assessing change. There is much discussion in the theoretical ecology literature of the relative merits of measures in terms of an underlying biological process model. However, such models seem so simplistic that we favour selection of suitable indices on the basis of what they measure, not on the basis of whether a suitable process model underlies them. For example, niche apportionment as envisaged by the geometric series model (Motomura 1932), in which the first species takes a fixed percentage of resources, then the next takes the same fixed percentage of what remains and so on, may provide a satisfactory fit for many assemblages but does not provide a realistic explanation of how the corresponding species distributions arose. If biological process models are required that have the potential to explain ecosystem composition and structure, then it seems necessary to model the biological processes that give rise to this composition and structure: that is, birth, survival and movement, and the effects of species interactions and the environment on these processes. Buckland *et al.* (2004) provide a state-space framework for modelling ecosystems in this way. Long time series of data from well-designed monitoring surveys allow such models to be fitted using computer-intensive techniques such as Markov chain Monte Carlo and sequential importance sampling (Newman *et al.* in preparation).

If biological process models are to be of any use to managers, they should model the major sources of uncertainty: observational error, demographic stochasticity and model uncertainty. If they do not, managers are unable to assess the risk associated with any management action. The above state-space framework allows all three major sources of uncertainty to be modelled (Buckland *et al.* 2004).

A major problem with realistic ecosystem models is that species interactions can be very complex and typical data from a monitoring scheme may be rather uninformative for fitting these interactions. Except for very simple ecosystems, substantial further studies might be required to identify how species interactions operate and to parametrize these interactions. Hence, a realistic ecosystem model, useful for aiding managers, will often be a long-term aspiration rather than a short-term goal.

We thank J. Nichols and an anonymous referee for particularly constructive and thoughtful reviews, which led to a much improved manuscript.

REFERENCES

- Alford, R. A., Dixon, P. M. & Pechmann, J. H. K. 2001 Global amphibian population declines. *Nature* **412**, 499–500.
- Balmford, A., Green, R. & Jenkins, M. 2003 Measuring the changing state of nature. *Trends Ecol. Evol.* **18**, 326–330.
- Buckland, S. T. Submitted. Point transect surveys for songbirds: robust methodologies.
- Buckland, S. T., Goudie, I. B. J. & Borchers, D. L. 2000 Wildlife population assessment: past developments and future directions. *Biometrics* **56**, 1–12.
- Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. & Thomas, L. 2001 *Introduction to distance sampling*. Oxford: Oxford University Press.
- Buckland, S. T., Newman, K. B., Thomas, L. & Koesters, N. B. 2004 State-space models for the dynamics of wild animal populations. *Ecol. Modell.* **171**, 157–175.
- Danielsen, F., Mendoza, M. M., Alviola, P., Balete, D. S., Enghoff, M., Poulsen, M. K. & Jensen, A. C. 2003 Biodiversity monitoring in developing countries: what are we trying to achieve? *Oryx* **37**, 407–409.
- Droege, S. 1990 The North American breeding bird survey. *Survey designs and statistical methods for the estimation of avian population trends* (ed. J. R. Sauer & S. Droege), pp. 1–4. Washington, DC: U.S.D.I. Fish Wildlife Service.
- Faith, D. P. 2002 Quantifying biodiversity: a phylogenetic perspective. *Conserv. Biol.* **16**, 248–252.
- Fewster, R. M., Buckland, S. T., Siriwardena, G. M., Baillie, S. R. & Wilson, J. D. 2000 Analysis of population trends for farmland birds using generalized additive models. *Ecology* **81**, 1970–1984.
- Freeman, S. N., Noble, D. G., Newson, S. E. & Baillie, S. R. 2003 Modelling bird population changes using data from the Common Birds Census and the Breeding Bird Survey. British Trust for Ornithology research report 303.
- Geissler, P. H. & Noon, B. R. 1981 Estimates of avian population trends from the North American breeding bird survey. *Stud. Avian Biol.* **6**, 42–51.
- Gotelli, N. J. & Colwell, R. K. 2001 Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* **4**, 379–391.
- Gregory, R. D., Noble, D., Field, R., Marchant, J., Raven, M. & Gibbons, G. W. 2003 Using birds as indicators of biodiversity. *Ornis Hungaria* **12–13**, 11–24.
- Hastie, T. J. & Tibshirani, R. J. 1990 *Generalized additive models*. London: Chapman and Hall.
- Houlahan, J. E., Findlay, C. S., Schmidt, B. R., Meyer, A. H. & Kuzmin, S. L. 2000 Quantitative evidence for global amphibian population declines. *Nature* **404**, 752–755.
- Houlahan, J. E., Findlay, C. S., Meyer, A. H., Kuzmin, S. L., Schmidt, B. R. *et al.* 2001 Reply to Alford *et al.* *Nature* **412**, 500.
- Kendall, W. L., Peterjohn, B. G. & Sauer, J. R. 1996 First-time observer effects in the North American breeding bird survey. *Auk* **113**, 823–829.
- Lande, R. 1996 Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos* **76**, 5–13.
- Link, W. A. & Sauer, J. R. 1997 Estimation of population trajectories from count data. *Biometrics* **53**, 488–497.
- Link, W. A. & Sauer, J. R. 1998 Estimating population change from count data: application to the North American breeding bird survey. *Ecol. Appl.* **8**, 258–268.
- Link, W. A. & Sauer, J. R. 1998 Estimating relative abundance from count data. *Austrian J. Stat.* **27**, 83–97.
- Magurran, A. E. 2004 *Measuring biological diversity*. Oxford: Blackwell Science.
- Magurran, A. E. & Phillip, D. A. T. 2001 Implications of species loss in freshwater fish assemblages. *Ecography* **24**, 645–650.
- May, R. M. 1975 Patterns of species abundance and diversity. *Ecology and evolution of communities* (ed. M. L. Cody & J. M. Diamond), pp. 81–120. Cambridge, MA: Harvard University Press.
- McGill, B. J. 2003 A test of the unified neutral theory of biodiversity. *Nature* **422**, 881–885.
- Moran, P. J. & Grant, T. R. 1991 Transference of marine fouling communities between polluted and unpolluted sites—impact on structure. *Environ. Pollut.* **72**, 89–102.
- Motomura, I. 1932 On the statistical treatment of communities. *Zool. Mag. Tokyo* **44**, 379–383. [In Japanese].
- Newman, K. B., Fernández, C., Buckland, S. T. & Thomas, L. In preparation. Inference for state-space models for wild animal populations.
- Pla, L. 2004 Bootstrap confidence intervals for the Shannon biodiversity index: a simulation study. *J. Agric. Biol. Environ. Stat.* **9**, 42–56.
- Pollard, E. & Yates, T. J. 1993 *Monitoring butterflies for ecology and conservation; the British Butterfly Monitoring Scheme*. London: Chapman and Hall.
- Pollock, K. H., Nichols, J. D., Simons, T. R., Farnsworth, G. L., Bailey, L. L. & Sauer, J. R. 2002 Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* **13**, 105–119.
- Reijnen, R., Foppen, R., ter Braak, C. & Thissen, J. 1995 The effects of car traffic on breeding bird populations in woodland. III. Reduction of density in relation to the proximity of roads. *J. Appl. Ecol.* **32**, 187–202.
- Rosenzweig, M. L. 1995 *Species diversity in space and time*. Cambridge: Cambridge University Press.
- Sauer, J. R., Peterjohn, B. G. & Link, W. A. 1994 Observer differences in the North American breeding bird survey. *Auk* **111**, 50–62.

- Seber, G. A. F. 1982 *The estimation of animal abundance and related parameters*, 2nd edn. New York: Macmillan.
- Simpson, E. H. 1949 Measurement of diversity. *Nature* **163**, 688.
- Siriwardena, G. M., Baillie, S. R., Buckland, S. T., Fewster, R. M., Marchant, J. H. & Wilson, J. D. 1998 Trends in the abundance of farmland birds: a quantitative comparison of smoothed Common Birds Census indices. *J. Appl. Ecol.* **35**, 24–43.
- Sørensen, L. L., Coddington, J. A. & Scharff, N. 2002 Inventorying and estimating subcanopy spider diversity using semiquantitative sampling methods in an Afromontane forest. *Environ. Entomol.* **31**, 319–330.
- Southwood, R. & Henderson, P. A. 2000 *Ecological methods*. Oxford: Blackwell Science.
- ter Braak, C. J.F., van Strien, A. J., Meijer, R. & Verstrael, T. J. 1994 Analysis of monitoring data with many missing values: which method?. (*Bird numbers 1992. Distribution, monitoring and ecological aspects* (ed. W. Hagemeyer & T. Verstrael) In *Proceedings of the 12th International Conference of the International Bird Census Committee and European Ornithological Atlas Committee*), pp. 663–673. The Netherlands: SOVON, Beek-Ubbergen.
- Thomas, L., Burnham, K. P. & Buckland, S. T. 2004 Temporal inferences from distance sampling surveys. In *Advanced distance sampling* (ed. S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers & L. Thomas), pp. 71–107. Oxford: Oxford University Press.
- Thompson, W. L., White, G. C. & Gowan, C. 1998 *Monitoring vertebrate populations*. San Diego: Academic Press.
- Warwick, R. M. & Clarke, K. R. 2001 Practical measures of marine biodiversity based on relatedness of species. *Oceanogr. Mar. Biol. A. Rev.* **39**, 207–231.
- Williamson, K. 1964 Bird census work in woodland. *Bird Study* **11**, 1–22.
- Yoccoz, N. G., Nichols, J. D. & Boulinier, T. 2001 Monitoring of biological diversity in space and time. *Trends Ecol. Evol.* **16**, 446–453.
- Yoccoz, N. G., Nichols, J. D. & Boulinier, T. 2003 Monitoring of biological diversity—a response to Danielsen *et al.* *Oryx* **37**, 410.

GLOSSARY

- BMS: British Butterfly Monitoring Scheme
 CBC: Common Birds Census
 NABBS: North American Breeding Bird Survey
 UK BBS: United Kingdom Breeding Bird Survey
 WBI: UK Wild Bird Index